Appendix A

Glossary of Terms

De-identified Data - A data set that has no, or limited, identifiers and for which a person with current knowledge of generally accepted scientific principles determines that the risk that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient, to identify an individual who is a subject of the information, has been reduced to the extent practicable. A graded approach must be used in balancing de-identification of the datasets and the usability of the dataset to accomplish the needed research.

Extraneous Data – Information in the dataset not relevant/needed for the specific research to be undertaken that may contain Personally Identifiable Information.

Human Subjects Research (HSR) – Any systematic investigation (including research development, testing, and evaluation) involving intervention or interaction with individuals or using their personally identifiable information or materials, designed to develop or contribute to generalizable knowledge. In addition to traditional biomedical and clinical studies, such research includes but is not limited to studies that –

(1) Use humans to examine devices, products or materials with the express purpose of investigating human-machine interfaces or evaluating environmental alterations when humans are the subjects being tested;
(2) Use personally identifiable bodily materials such as cells, blood, tissues, urine, or hair, even if the materials were collected previously for a purpose other than the current research;
(3) Collect and use personally identifiable information such as genetic information or medical and exposure records, even if the information was collected previously for a purpose other than the current research;
(4) Collect personally identifiable or non-identifiable data, surveys, or questionnaires through direct intervention or interaction with individuals; and
(5) Search for generalizable knowledge about categories or classes of subjects (e.g., linking job conditions of worker populations to hazardous or adverse health outcomes).

Human Terrain Mapping (HTM) - Research and data gathering activities primarily conducted for military or intelligence purposes to understand the “human terrain”— the social, ethnographic, cultural, and political elements of the people among whom the U.S. Armed Forces are operating and/or in countries prone to political instability. This work includes observations, questionnaires, and interviews of groups of individuals, as well as modeling and analysis of collected data, and may become the basis for U.S. military actions in such locations. In addition to Human Terrain Mapping (HTM), such activities are often referred to as human social culture behavior (HSCB) and human terrain systems (HTS) studies. It is DOE policy that HTM activities will be managed as HSR.

HTM Data - Data collected or used as part of HTM efforts, as described above, as well as any auxiliary data on the same group(s) of individuals.

Identifier – See Appendix B.

Institutional Review Board (IRB) - A committee or board established by an institution that performs initial and continuing reviews of research involving human subjects, and is registered with the Office for Human Research Protections (OHRP) and designated on a Federal Wide Assurance (FWA).

Merged Data - Data from two or more datasets that has been combined into a single new data set.

Personally Identifiable Information - Any information collected or maintained about an individual, including but not limited to, education, financial transactions, medical history and criminal or employment history, and information that can be used to distinguish or trace an individual's identity, such as his/her name, Social Security number, date and place of birth, mother's maiden name, biometric data, and any other personal information that is linked or linkable to a specific individual. Refer to DOE O 206.1, *Department of Energy Privacy Program*.

Appendix B

Identifiers, Quasi-Identifiers and Data Security Requirements

Introduction

The purpose of this appendix is to provide a reference for IRB members and investigators on:

1) Identifiers: data that can uniquely identify individuals
2) Quasi-identifiers: data that does not explicitly identify individuals but when used in combination with other data can do so
3) Potential data security requirements: Requirements the IRB can ask to be enforced on data.

Identifiers

1. HIPAA “Safe Harbor” Fields

As of February 2, 2011, the Health Insurance Portability and Accountability Act of 1996 (HIPAA), as amended. Privacy and Security Rules indicates the following data be removed or de-identified before sharing medical data on individuals. Note that for HTM projects, some of these identifiers may not apply and are listed here as a reference.

Data Field	**Notes**
Names	
Geographic subdivisions	Any geographic subdivision less than a state. *
Dates (except year)	Includes birth date, admission date, discharge date, date of death, and ages > 89. **
Telephone /Fax numbers	
Email	
Social Security Numbers	
Medical Record numbers	
Health plan beneficiary numbers	

Account numbers	
Certificate/license numbers	
Vehicle identifiers (including license plate numbers)	
Device identifiers and serial numbers	
URLs	
IP Addresses	
Biometric identifiers (including finger and voice prints)	
Full face photographic images	

- *The initial three digits of the zipcode are allowed in certain cases. See 45 CFR §164.514 for more details.
- ** Ages above 89 can be aggregated into a >90 category.

Quasi-Identifiers

The following table contains quasi-identifiers sets that have been used to re-identify data sets. Each set has a reference that provides more detail. Re-identification based on quasi-identifier sets is dependent upon many variables (such as the existence and availability of an auxiliary data set) and thus this list should be taken primarily as a starting point for further discussion.

Quasi-Identifier sets	**Note**	**Reference**
Zip code, birth date, sex	References: Indicate that >50% of U.S. individuals have a unique combination of these fields.	Sweeney, 2000
Movie title, rating, date of movie ratings	Re-identification of some Netflix users based on these fields, using data from the public movie rating site IMDB.com	Narayanan, 2008
International Statistical Classification of Diseases and Related Health Problems (ICD-9)	Set of diagnosis codes for a patient can possibly be unique.	Loukides, 2010

Potential Data Security Requirements

Some data security requirements that the IRB may impose on a project as a function of the data properties (e.g., level of de-identification, size of data set) and project properties (e.g., utility of certain attributes) include:

- Data use only on systems disconnected from the network;
- When not in use, the dataset must be encrypted according to lab approved encryption program and/or on a storage device not physically connected to a machine; and
- Dataset must be destroyed or turned over to the sponsor at the end of the project.

Case Studies

See the following link for case studies of datasets that included quasi-identifiers and the actions that subsequently took place. (*Link will be inserted once case studies have been developed*)

References

L. Sweeney. Uniqueness of Simple Demographics in the U.S. Population. Technical Report LIDAP-WP4, Laboratory for International Data Privacy, 2000.

A. Narayanan and V. Shmatikov. Robust De-Anonymization of Large Sparse Datasets. In Proceedings of the 29th IEEE Symposium on Security and Privacy, 2008.

G. Loukides, J. C. Denny, and B. Malin. The Disclosure of Diagnosis Codes Can Breach Research Participants Privacy. Journal of the American Medical Informatics Association, 17:3.

Health Insurance Portability and Accountability Act of 1996 (HIPAA), as amended.

Appendix C

HTM Data Security Agreement

This HTM Data Security Agreement is entered on, 201between

______________________________________ ______________________________ ("IRB") and
IRB Chair (please print) *DOE Laboratory/Site*

______________________________________ ______________________________ ("PI").
Principal Investigator (please print) *Organization*

This Agreement establishes the terms and conditions under which the PI will protect the HTM data

____________________________________ for the ____________________________________
HTM dataset (s) name *HTM research project name*

as approved for use by the IRB (see attached HTM Data Checklist). If a Data Security Agreement exists for the use of this HTM data between a Sponsor and the PI, this Agreement must complement and not contradict those terms and conditions.

Use of this HTM data, in whole or in part, for other HTM and/or human subjects research projects will be subject to prior approval by the IRB.

The terms and conditions of this Agreement are developed jointly between the PI and IRB during the HTM Data Checklist review. This Agreement can be changed only by a written modification of the agreement by the party signatories (or their replacements) to this Agreement or by the parties adopting a new agreement in place of this Agreement.

The PI will be designated as custodian of this HTM data and will be responsible for complying with all conditions of use and for establishment and maintenance of security arrangements as specified in this Agreement to prevent unauthorized use and disclosure of this HTM data.

Access to the HTM data will be limited to the minimum number of individuals who need access to this data. Describe below how access will be limited and/or controlled.

The appropriate administrative, technical, and physical safeguards will be established to prevent unauthorized use or access to this HTM data. Describe below what safeguards will be used to store, distribute, transmit, publish, etc. the HTM data.

Upon project completion, the HTM data will be archived/transferred/destroyed/etc. as described below, and the IRB notified once these steps have been completed.

Derivative data created from this HTM data will be managed in the same manner as the original data, unless otherwise approved by the IRB.

The PI will inform the project team (to include team members, subcontractors, collaborators, and any other parties with access to the HTM dataset) of this HTM Data Security Agreement and will have them countersign below to document their awareness and expected compliance to these HTM Data security terms and conditions.

The PI agrees to notify the IRB immediately if the Agreement or any provision of this Agreement has been breached. Such notification will include the identity of such individuals and the nature of the breach.

The PI agrees to notify the IRB and their management if the HTM Data has been lost, stolen, or compromised in any way. The response is expected to be in alignment with security incident reporting and the Sponsor of the HTM Dataset will be notified as appropriate.

This Agreement may be terminated by either party at any time for any reason upon ten (10) days written notice. Upon such notice, the HTM Data will be handled per the project completion plan described above.

This Agreement expires upon written notification by the PI to the IRB that the project has been completed and data properly disposed and/or secured (as described above). This Agreement will be transferred to and signed by a new PI after any change of custodianship.

____

PI Signature

____

Printed Name

____

Date

____

IRB Chair Signature

____

Printed Name

____

Date